Check 34. Optimizing Field-Aware Factorization Machine.pdf

by Arta Sundjaja

Submission date: 07-May-2019 05:24PM (UTC+0700)
Submission ID: 1102510027
File name: 34._Optimizing_Field-Aware_Factorization_Machine.pdf (4.09M)
Word count: 2348
Character count: 10863

2018 3rd International Conference on Computer and Communication Systems

Optimizing Field-Aware Factorization Machine with Particle Swarm Optimization on Online Ads Click-through Rate Prediction

Michael Reynaldo Phangtriastu
School of Computer Science
BINUS University
Jakarta, Indonesia
e-mail: mphangtriastu@binus.edu

Sani Muhamad Isa
School of Computer Science
BINUS University
Jakarta, Indonesia
e-mail: sani.m.isa@binus.ac.id

***Abstract*—Online advertising industry is grow larger along with the increasing numbers of internet users. To make ads industry to be more efficient, prediction model for ads' click-through rate is needed. In this research, Field-aware Factorization Machine (FFM) is going to be optimized using Particle Swarm Optimization (PSO) on FFM parameters to increase the accuracy of the FFM. In this research, FFM and PSO-FFM is compared with accuracy and execution time. Our experimental results show PSO can increase FFM performance.**

Keywords-click-through rate; FFM; PSO-FFM

I. INTRODUCTION

Online advertising is becoming more popular than conventional advertising, such as newspaper or television. Counting return on investment (ROI) of online advertising more possible because all factor and data can be tracked [1]. For the payment, advertiser has varied options to pay. Advertiser can pay per view, per click, etc. [2]. Current business is needed to using online advertising to stay competitive.

One of online advertising factor is click-through rate (CTR). CTR is a percentage calculated by dividing click counts from ad with view counts from ad. This factor is crucial to look how efficient an ad being placed to targeted users [1] [2].

CTR prediction is currently plays important role in advertising industry [2] [3]. With predicting CTR, advertiser can count the most optimal ROI before placing ads. In CTR prediction problems, prediction model need to predict is an ad being clicked or not from a click log. The click logs are generated from users of a website. The user criteria and behavior are collected by advertising company. The user criteria can be user location, browser used, the type of mobile phone used, the operating system used, and so on. The behavior can be click coordinate, click frequency, scroll behavior, and so on. For an example, click coordinates can be used to generate an information like click heatmap in Figure 1.

Because of the need of good CTR prediction system, there are many CTR prediction competition being held. The most popular algorithm that win many competitions is Field-aware Factorization Machine (FFM). FFM success to outperforms existing models, such as logistic regression, SVM, etc. [3]. The FFM is variation from Factorization Machine (FM) [4].

Figure 1. Click heatmap example.

In this paper, FFM model is going to be optimized with an optimization algorithm, Particle Swarm Optimization (PSO). The PSO is one of evolutionary-based optimization algorithm with good performance. Evolutionary-based optimization algorithms are the algorithm that mimicking nature phenomenon. There are some researches that comparing PSO with another evolutionary-based optimization algorithm [5]. From the result of the previous work, PSO relatively has good result compared to other algorithms. In this research, the FFM parameters model (k, λ, η) are going to be optimized using PSO.

II. FACTORIZATION MACHINE (FM)

The FFM is variate FM with adding more dimensions to FM latent vector [3]. The FM formula [4]:

$$y(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \left(\sum_{f=1}^{k} v_{i,f} \cdot v_{j,f}\right) x_i x_j$$

Where *y(x)* is the predicted class, *w* are weight, *x* are value of feature, and *v* are latent vectors. *n* is total features and *k* is number of latent vectors. The *w* will be updated with gradient descent [4]:

$$\frac{\delta}{\delta\theta}y(x) = \begin{cases} 1 & if\ \theta\ is\ w_0 \\ x_i & if\ \theta\ is\ w_1 \\ x_i \sum_{j=1}^{n} v_{j,f}x_j - v_{j,f}x_i^2 & if\ \theta\ is\ v_{i,f} \end{cases}$$

III. FIELD-AWARE FACTORIZATION MACHINE (FFM)

In FFM, *y(x)* formula from FFM is becoming:

$$y(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n}\sum_{j=i+1}^{n}(\sum_{f=1}^{k} v_{i,j,f} \cdot v_{j,i,f})x_i x_j$$

As you can see, the latent vector *k* is becoming three-dimensional.

The FFM optimization problem is [3]:

$$\min_{w} \frac{\lambda}{2}\|w\|_2^2 + \sum_{i=1}^{m} \log(1 + \exp(-y_i\, y_i(x)))$$

Where λ is regularization parameter, *m* are size of data, *y* are the data label.

To update the weights, first *g* value need to be counted for every *w*:

$$g_{i,j} \equiv \nabla w_{i,j} f(w) = \lambda . w_{i,j} + \frac{-y}{1 + \exp(y\, y(x))} . w_{i,j} x_i x_j$$

After counting *g* value, g^2 added to *G*, accumulative of *g*.

$$G_{i,j} \leftarrow G_{i,j} + g_{i,j}^2$$

G and current *g* is used for updating *w*, η is learning rate parameter:

$$w_{i,j} \leftarrow w_{i,j} - \frac{\eta}{\sqrt{G_{i,j}}} g_{i,j}$$

At the beginning of FFM, *w* value is set to uniform random number [0,1/*k*] and *G* is set to one [3].

All input data for FM and FFM are sparse matrix. The standard input for FFM is:

label field1:feat1:val1 field2:feat2:val2

Figure 2. FFM input format.

In criteo dataset, the first 13 column has numeric value. There are two methods for inputting numeric data to FFM algorithm [3].

The first method is using the number as the value. For example, data from Figure 3 will be: Yes AR:AR:45.73 Hidx:Hidx:2 Cite:Cite:3.

The second method is using the number as the field name, the value is 1. In this case, the number is treated as categorical data. For example, data from Figure 3 will be: Yes AR:45:1 Hidx:2:1 Cite:3:1.

Accepted	AR	Hidx	Cite
Yes	45.73	2	3
No	1.04	100	50,000

Figure 3. Numeric data example.

IV. PARTICLE SWARM OPTIMIZATION (PSO)

PSO are is one of evolutionary algorithm that inspired by social behavior of a flock of migrating birds trying to reach a good place to live, but still unknown. In PSO, each solution is a 'bird' referred as a 'particle' [5]. The particle has its own position and velocity. Every iteration, velocity will be updated according to current fitness, personal best fitness, and global best fitness, and position will be updated according to current velocity. Every particle kept its personal best fitness. Global best is the best fitness that flock has so far. To update position, PSO using [6]:

$$x_i(t+1) = x_i(t) + v_i(t+1)$$

Where *x* are positions and *v* are velocities. *t* is current iteration.

To updating velocity, PSO has many variations, one of them is called global best PSO (gbest-PSO), using [6]:

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_1(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_2(t)[\tilde{y}_j(t) - x_{ij}(t)]$$

Where y are personal best for each particle in each features and $\tilde{y}_j$ are global best from all particle in each features. *c* are constant values that being since the start of PSO, and *r* are random values that changed every iteration.

In this research, the PSO algorithm is used for optimizing FFM parameters. The boundaries of parameters are:

TABLE I. OPTIMIZING PARAMETER BOUNDARIES

Parameter	Lower bound	Upper bound
k	1	20
η	0.01	0.5
λ	0.00001	0.01

The boundaries are being set based on the previous work that using $\eta = 0.2, \lambda = 2\ x\ 10^{-5}, k = 4$.

V. DATA PREPROCESSING

The criteo data has imbalanced between true labeled data and false labeled data. For day 1 data, true labeled is about 3% and false labeled data is about 97%.

In day 1 dataset, there are some missing values in a row of data. Only about 21% data has complete data.

TABLE II. COUNT OF LABEL OF DAY 1 CRITEO DATASET

Rows	True labeled	False labeled
199.563.535	6.373.290	193.190.245
Percentage	3.19%	96.81%

TABLE III. COUNT OF MISSING VALUE OF DAY 1 CRITEO DATASET

Rows	Complete	Missing
199.563.535	42.451.134	157.112.401
Percentage	21.27%	78.73%

In this research, the true labeled data and false labeled data is rebalanced at the sampling process. The size of true labeled data is minimal of 40% of the sample dataset. The incomplete row will not be using as sample dataset. Only complete row will be added to sample dataset.

In the dataset, there are some values that rarely appear. We are going to do an experiment to compress the rare value to be one same value. We assume this technique can increase performance because the FFM model is using matrix. With larger variety of value, the size of matrix is become larger. By eliminating some variety of value, it can decrease the execution time.

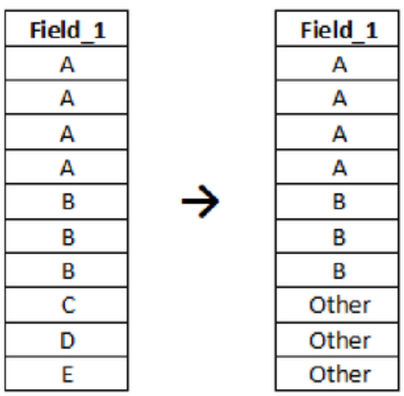

Figure 4. Example of transforming rare value.

We are using a percentage to be a threshold to determine a value as a rare value, later will be called rarity percentage. If the count of the value is less than or equal to the threshold, the value is changed as other value. For an example, on the Figure 4, there are 10 data. If the rarity percentage is 20%, the value that has count below than 2 is changed to 'other'. The 'A' value has 40% of dataset, 'B' value has 30% of dataset, while 'C', 'D', and 'E' value has 10% of dataset. In this case, 'C', 'D', and 'E' value changed to 'other' values.

VI. RESULTS AND DISCUSSIONS

In this research, dataset that used is day 1 data from criteo click log dataset. Sample variations are: 100 rows, 1.000 rows, and 10.000 rows. For each sample, the variations of rarity percentage are: 0%, 1%, 2%, and 3%.

TABLE IV. FFM AND PSO-FFM ACCURACY RESULTS

Rows	Rarity Percentage	FFM	PSO - FFM
100	0	58.06	61.29
100	1	74.19	74.19
100	2	67.74	77.42
100	3	64.52	74.19
1000	0	73.23	80.00
1000	1	76.05	79.61
1000	2	74.76	80.67
1000	3	76.38	80.58
10000	0	77.76	80.70
10000	1	77.09	80.70
10000	2	76.89	80.70
10000	3	77.30	80.34

TABLE V. FFM AND PSO-FFM EXECUTION TIME (MS)

Rows	Rarity Percentage	FFM	PSO - FFM
100	0	12752	161721
100	1	5584	96852
100	2	5450	90637
100	3	5353	94187
1000	0	13747	285229
1000	1	6848	162323
1000	2	6780	162108
1000	3	6621	136166
10000	0	29498	482930
10000	1	19769	331740
10000	2	19723	322823
10000	3	18963	322369

TABLE VI. AVERAGE OF FFM AND PSO-FFM ACCURACY RESULTS BASED ON RARITY PERCENTAGE

Rarity Percentage	FFM	PSO - FFM
0	69.68	74.00
1	75.78	78.17
2	73.13	79.60
3	72.73	78.37

From the result above (Table 4 and Table 5), can be seen the result of every sample dataset with every variation of rarity percentage. By increasing the size of dataset, it can affect the accuracy. On the other hand, using rarity percentage can affect the accuracy and execution time. The result will be average based on rarity percentage and based on size of sample to give more look.

With using rarity percentage, the accuracy and execution time is increased significantly. In FFM, the accuracy of 0% rarity percentage is about 70%, otherwise the accuracy of 1% rarity percentage is about 76%. In PSO-FFM, the accuracy of 0% rarity percentage is about 74%, otherwise the accuracy of 1% rarity percentage is about 78%.

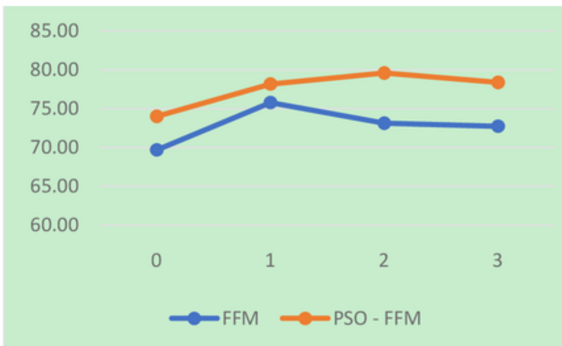

Figure 5. Average of FFM and PSO-FFM accuracy results based on rarity percentage

Based on the execution time, in FFM the 0% rarity percentage run in 18 seconds, otherwise the 1% rarity percentage is run in 10 seconds. In PSO-FFM the 0% rarity percentage run in 30 seconds, otherwise the 1% rarity percentage run in 19 seconds.

Based on the experiment, increasing the rarity percentage (1%, 2%, and 3%) does not make significant different to the performance.

TABLE VII. AVERAGE OF FFM AND PSO-FFM EXECUTION TIME (MS) BASED ON RARITY PERCENTAGE

Rarity Percentage	FFM	PSO - FFM
0	18666	309960
1	10734	196972
2	10651	191856
3	10312	184241

TABLE VIII. TVERAGE OF FFM AND PSO-FFM ACCURACY RESULTS BASED ON NUMBER OF ROWS

Rows	FFM	PSO – FFM
100	66.13	71.77
1000	75.10	80.22
10000	77.26	80.61

As we can see, the accuracy of in PSO – FFM is increase for about 5% for each data sample compared to FFM results. Increasing dataset size is also increase accuracy.

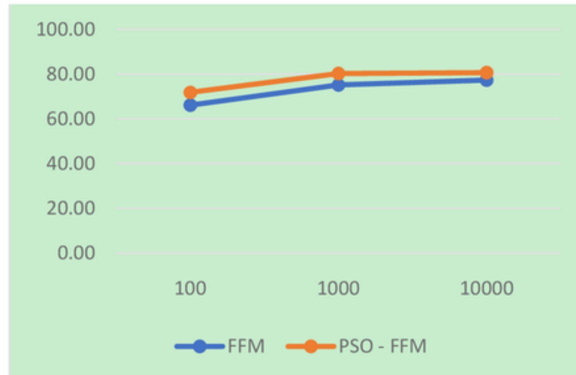

Figure 6. Average of FFM and PSO-FFM accuracy results based on number of rows

TABLE IX. AVERAGE OF FFM AND PSO-FFM EXECUTION TIME (MS) BASED ON NUMBER OF ROWS

Rows	FFM	PSO - FFM
100	7285	110849
1000	8499	186457
10000	21988	364966

For the execution time, PSO-FFM has longer time needed to be executed. The difference is about 17 times slower.

VII. CONCLUSION AND FUTURE WORK

From this research, it can be concluded that PSO is able to increase the accuracy of FFM, but for the trade of, the execution time is increased. On the other hand, using rarity percentage can decrease execution time and increase the accuracy both in FFM and PSO-FFM.

For the future work, PSO-FFM can be optimized by doing parallel implementation to reduce the execution time, comparing the PSO with other optimizing technique for looking the most suitable optimizing technique for FFM. Optimizing rarity percentage is also interesting topic to cover.

REFERENCES

[1] L. Harden and B. Heyman, Marketing by the numbers: how to measure and improve the ROI of any campaign, AMACOM Div American Mgmt Assn, 2011.

[2] O. Chapelle, E. Manavoglu and R. Rosales, "Simple and scalable response prediction for display advertising," *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2015.

[3] Y. Juan, Y. Zhuang, W.-S. Chin and C.-J. Lin, "Field-aware factorization machines for CTR prediction," in *Proceedings of the 10th ACM Conference on Recommender Systems*, 2016.

[4] S. Rendle, "Factorization Machines," in *2010 IEEE 10th International Conference on Data Mining (ICDM)*, 2010.

[5] E. Elbeltagi, T. Hegazy and D. Grierson, "Comparison among five evolutionary-based optimization algorithms," *Advanced engineering informatics,* 2005.

[6] A. P. Engelbrecht, Computational intelligence: an introduction, John Wiley & Sons, 2007.

Check 34. Optimizing Field-Aware Factorization Machine.pdf

ORIGINALITY REPORT

6%	2%	7%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Ricky Anderson, Aryo Pradipta Gema, Suharjito, Sani M. Isa. "Facial Attractiveness Classification using Deep Learning", 2018 Indonesian Association for Pattern Recognition International Conference (INAPR), 2018 Publication	2%
2	www.csie.ntu.edu.tw Internet Source	1%
3	"Table of contents", 2018 3rd International Conference on Computer and Communication Systems (ICCCS), 2018 Publication	1%
4	www.spie.org Internet Source	1%
5	Tri Fennia Lesmana, Sani Muhamad Isa, Nico Surantha. "Sleep Stage Identification Using the Combination of ELM and PSO Based on ECG Signal and HRV", 2018 3rd International Conference on Computer and Communication Systems (ICCCS), 2018	1%

Publication

6	Lecture Notes in Computer Science, 2013. Publication	1%
7	Elbeltagi, Emad E.. "Swarm Intelligence for Large-Scale Optimization in Construction Management", Metaheuristic Applications in Structures and Infrastructures, 2013. Publication	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		

Check 34. Optimizing Field-Aware Factorization Machine.pdf

GRADEMARK REPORT

FINAL GRADE

/0

GENERAL COMMENTS

Instructor

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5